

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/200202>

Please be advised that this information was generated on 2019-12-04 and may be subject to change.

Sociale Kunstmatige Intelligentie

INAUGURELE REDE DOOR PROF. DR. TIBOR BOSSE

•
in au
gurele
redo

change perspective

Radboud Universiteit



INAUGURELE REDE
PROF. DR. TIBOR BOSSE



Recente ontwikkelingen binnen de Kunstmatige Intelligentie (AI) hebben een enorme impact op onze samenleving, en die impact zal de komende jaren alleen maar groeien. Om te garanderen dat deze ontwikkelingen het menselijk welzijn positief beïnvloeden, is een mensgerichte, sociale focus

op AI cruciaal. Moderne AI-systemen moeten niet alleen beslissingen kunnen nemen, maar ook kunnen interacteren en samenwerken met mensen. Voorbeelden van zulke sociale AI-systemen zijn sociale robots, digitale assistenten en chatbots. Om dergelijke systemen optimaal te benutten is er niet alleen behoefte aan geavanceerde technologie, maar ook aan gedegen kennis van de effecten van technologie op menselijk gedrag. Dit vereist multi-disciplinair onderzoek op het grensvlak van de Informatica en de Sociale Wetenschappen. Onderzoek naar betere algoritmen op gebieden als natuurlijke taalverwerking en emotieherkenning dient te worden gecombineerd met gedragswetenschappelijk onderzoek, zoals gedaan binnen de mediapsychologie.

Tibor Bosse studeerde Computer Science aan de Vrije Universiteit Amsterdam met als specialisatie Artificial Intelligence. In 2005 promoveerde hij aan dezelfde universiteit op het proefschrift *Analysis of the Dynamics of Cognitive Processes*, waarin hij een nieuwe methode introduceerde voor het ontwikkelen van computermodellen van cognitieve processen. Bosse werkte verder als postdoctoraal onderzoeker aan de Vrije Universiteit, en als gast-onderzoeker aan de Universiteit van Amsterdam, de Technische Universiteit Delft en TNO Soesterberg.

Radboud Universiteit



SOCIALE KUNSTMATIGE INTELLIGENTIE

Sociale Kunstmatige Intelligentie

Rede uitgesproken bij de aanvaarding van het ambt van hoogleraar Communication Science and Artificial Intelligence aan de Faculteit der Sociale Wetenschappen van de Radboud Universiteit op vrijdag 18 januari 2019

door prof. dr. Tibor Bosse

Opmaak en productie: Radboud Universiteit, Facilitair Bedrijf, Post & Print
Fotografie omslag: Bert Beelen

© Prof. dr. Tibor Bosse, Nijmegen, 2019

Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar worden gemaakt middels druk, fotokopie, microfilm, geluidsband of op welke andere wijze dan ook, zonder voorafgaande schriftelijke toestemming van de copyrighthouder.

*Mijnheer de rector magnificus,
geachte leden van het college van bestuur,
zeer gewaardeerde toehoorders,*

Vraagt u zich wel eens af hoe de wereld er over tien jaar uitziet? Of over vijftig jaar? Helaas zullen velen van ons die tijd niet meer meemaken, maar voor wie zich bekommert om het lot van de generaties na ons, is dat een uiterst relevante vraag. Het is ook een hele moeilijke vraag. ‘De wereld verandert tegenwoordig razendsnel’ is een cliché dat je de laatste tijd overal hoort, of het nu op een feestje is, tijdens een lezing, of in de media. Maar het is meer dan een cliché. Het is een constatering die verregaande gevolgen heeft. Zo is het, in vergelijking met vroeger, veel belangrijker geworden om na te denken over de impact van technologie op ons dagelijks leven.

Het extreme tempo waarin de wereld momenteel verandert wordt treffend geïllustreerd door Yuval Noah Harari, auteur van bestsellers als *Sapiens* en *Homo Deus*. In zijn nieuwste boek *21 lessen voor de 21ste eeuw* zegt hij het volgende: “A thousand years ago, in 1018, there were many things people didn’t know about the future, but they were nevertheless convinced that the basic features of human society were not going to change. If you lived in China in 1018, you knew that by 1050 the Song Empire might collapse, the Khitans might invade from the north, and plagues might kill millions. However, it was clear to you that even in 1050 most people would still work as farmers and weavers, rulers would still rely on humans to staff their armies and bureaucracies, men would still dominate women, life expectancy would still be about 40, and the human body would be exactly the same¹.”

Hoe anders is de situatie anno 2019. Als we vandaag een voorspelling zouden moeten doen over de wereld in 2050 dan zouden we geen idee hebben. Zouden we nog steeds op vrijdagmiddag op pad gaan om een inaugurele rede bij te wonen - wellicht per zelfrijdende auto? Of zouden we zulke redes gewoon van huis uit volgen, op onze smartphone, via Virtual Reality, of middels een totaal nieuw medium? Of zouden oraties al lang niet meer gehouden worden, omdat het vak van hoogleraar is overgenomen door robots? We weten het niet. Het enige dat echt zeker is, is dat de wereld constant zal veranderen.

In mijn redevoering zal ik betogen dat er in deze veranderende wereld een sterke behoefte is aan synergie tussen sociale en technische wetenschappen, en in het bijzonder tussen de communicatiewetenschap en de kunstmatige intelligentie. Ik zal beginnen met een beschrijving van technologische ontwikkelingen, en zal gaandeweg overgaan op de rol van de mens en de communicatiewetenschap.

KUNSTMATIGE INTELLIGENTIE

Als we de meest invloedrijke veranderingen van de afgelopen decennia in ogenschouw nemen, dan spelen de ontwikkelingen binnen de digitale technologie een belangrijke rol. In 1936 presenteerde de Britse wiskundige Alan Turing het idee van de *universele rekenmachine*, wat de basis vormde voor de moderne computer². Sinds dat moment is er binnen de informatica grote vooruitgang geboekt, waardoor computers in minder dan een eeuw tijd enorme stappen hebben gemaakt wat betreft eigenschappen als rekenkracht, opslagcapaciteit en geheugen. Informaticus Michael Wooldridge beschrijft vijf aanhoudende trends binnen de geschiedenis van de digitale technologie, te weten *ubiquity* (computers zijn overal aanwezig), *interconnection* (computersystemen raken steeds meer met elkaar verbonden), *intelligence* (computers voeren steeds complexere taken uit), *delegation* (we geven steeds meer controle aan computers) en *human-orientation* (onze communicatie met computers wordt steeds intuïtiever)³. Het gevolg van deze ontwikkelingen is dat we steeds meer te maken krijgen met computersystemen die zelfstandig taken voor ons uitvoeren. Dit heeft geleid tot de opkomst van een nieuw deelgebied van de informatica, dat we tegenwoordig kennen als kunstmatige of artificiële intelligentie (AI).

Kunstmatige intelligentie is een vakgebied dat zich richt op het onderzoeken en ontwikkelen van intelligente systemen. Hoewel de term al in 1956 werd geïntroduceerd door John McCarthy⁴, heeft AI pas sinds een aantal jaren de aandacht van het grote publiek. Je kunt tegenwoordig geen artikel of documentaire meer bekijken zonder te worden overspoeld met voorspellingen over hoe AI de wereld gaat veranderen. Zo heeft Vladimir Poetin onlangs geroepen dat de natie die de leider binnen de AI wordt zal heersen over de wereld⁵. En de topman van Google, Sundar Pichai, zei afgelopen jaar dat AI meer impact zal hebben dan elektriciteit en vuur⁶. Jammer genoeg geven dergelijke uitspraken in de media vaak aanleiding tot het trekken van extreme conclusies, met aan de ene kant van het spectrum de stelling dat AI al onze problemen zal oplossen en aan de andere kant dat AI de mensheid zal vernietigen.

De werkelijkheid ligt uiteraard genuanceerder, maar desondanks zit er een kern van waarheid in de uitspraken van Poetin en Pichai. De meeste experts zijn het er tegenwoordig wel over eens dat AI een zogenaamde *disruptive technology* is geworden, een technologie die vrijwel alle facetten van onze maatschappij fundamenteel zal innoveren. Ook nu zijn er al voorbeelden te over van AI-toepassingen die wij in ons dagelijks leven gebruiken, waaronder smartphone assistenten zoals Apple's Siri, gepersonaliseerde aanbevelingen van mediabedrijven zoals YouTube en navigatiediensten zoals Google Maps.

Het succes van dit soort toepassingen is vooral te danken aan recente doorbraken binnen de machine learning, een belangrijk deelgebied van de AI. De term machine learning verwijst naar algoritmen die machines in staat stellen om hun prestatie binnen bepaalde taken te verbeteren door te leren van grote hoeveelheden voorbeelddata. Hierdoor kunnen machine-learningapplicaties 'intelligent' gedrag vertonen zonder gebruik te maken van expliciet geformuleerde regels, wat in het verleden de gangbare manier was om AI-systemen te ontwikkelen. Binnen de machine learning zijn de ontwikkelingen de laatste jaren in een stroomversnelling geraakt met dank aan een samenspel van factoren: allereerst een toename in rekenkracht, waardoor AI-systemen hun berekeningen sneller kunnen uitvoeren. Een tweede belangrijke factor is de exponentiële groei in de hoeveelheid gegevens die digitaal beschikbaar zijn, resulterend in het tijdperk van *big data*. We kunnen tegenwoordig alles meten, variërend van ons zoekgedrag op Internet tot het aantal stappen dat we zetten, en we hebben ook nog eens de capaciteit om veel van die data op te slaan. En ten derde zijn de algoritmen zelf sterk verbeterd. In het bijzonder was een belangrijke doorbraak de opkomst van *deep learning*, een specifieke variant van machine learning die losjes is geïnspireerd door de werking van het menselijk brein⁷.

Met dank aan technieken als deep learning zijn er de afgelopen jaren veel punten toegevoegd aan de al lange lijst van taken waarin machines beter zijn dan mensen. Een bekend voorbeeld is het spelen van het Chinese bordspel *Go*. Hoewel de wereldkampioen schaken al in 1997 werd verslagen door een computer⁸, hielden de experts het lange tijd voor onmogelijk dat iets vergelijkbaars bij *Go* zou gebeuren, omdat dit spel nog vele malen complexer is. De wereld keek dan ook verbouwereerd toe toen in 2016 de regerend wereldkampioen *Go* zijn meerdere moest erkennen in het door Google DeepMind ontwikkelde programma *AlphaGo*⁹. Maar niet alleen het spelen van spelletjes kunnen computers beter dan mensen. Ook wanneer het gaat om het interpreteren van de omgeving kunnen AI-systemen inmiddels indrukwekkende prestaties leveren. Zo bestaan er bijvoorbeeld al AI-systemen die menselijke prestaties overtreffen bij taken als liplezen¹⁰, het transcriberen van gesproken teksten¹¹, en het herkennen van longkanker op basis van röntgenfoto's¹². Doordat de omvang van de beschikbare data wereldwijd alleen maar zal toenemen en er voor algoritmen dus nog meer input beschikbaar komt om van te leren, is het niet vreemd om te veronderstellen dat er de komende tijd nog talloze voorbeelden aan dit lijstje zullen worden toegevoegd.

BEPERKINGEN VAN HUIDIGE AI

Toch moeten we niet de vergissing maken om te denken dat de huidige generatie van AI-technieken binnenkort elk gewenst probleem kan oplossen. Een kenmerk van de geschiedenis van de AI is dat successen en teleurstellingen elkaar afwisselen. En hoewel het verleidelijk is om de snelle ontwikkelingen van de afgelopen tien jaar te extrapolere-

ren naar een toekomstbeeld waarop AI de mens op alle fronten voorbij zal streven, hebben ook de nieuwste AI-technieken bepaalde tekortkomingen. De belangrijkste beperking is wellicht dat alle bestaande toepassingen zijn gericht op vrij specifieke taken. Zo kan een algoritme dat is ontworpen om röntgenfoto's te analyseren op dat vlak fantastische resultaten bereiken, maar ditzelfde algoritme zal volstrekt onbruikbaar zijn voor bijvoorbeeld het vertalen van teksten. Deze vorm van AI, die zich richt op een specifiek, afgebakend domein, wordt ook wel *weak AI* genoemd. En hoewel er ontelbaar veel nuttige toepassingen bestaan op basis van *weak AI*, is dat een fundamenteel ander concept dan de AI die we vaak in science fiction films zien. Daar gaat het doorgaans om *strong AI*: machines met bewustzijn en gevoelens, die bovendien in staat zijn om hun intelligentie op ieder gewenst probleem toe te passen. Dat punt heeft het huidige AI-onderzoek nog lang niet bereikt¹³: om machines met bewustzijn te kunnen ontwikkelen weten we nog veel te weinig over wat bewustzijn precies inhoudt, en om AI te ontwikkelen die generiek is - dus volledig onafhankelijk van een specifiek domein - is waarschijnlijk een fundamentele verbetering nodig ten opzichte van bestaande technieken.

Een tweede belangrijke beperking van *state-of-the-art AI*, die bij uitstek van toepassing is op de nieuwste machine-learningtechnieken, betreft het geven van uitleg aan de gebruiker. Systemen die gebruik maken van machine learning hebben doorgaans namelijk het karakter van een *black box*: ze worden gevoed met bepaalde input, en op basis daarvan wordt er output gegenereerd in de vorm van een beslissing. Deze uitkomsten zijn vaak uitstekend van kwaliteit, maar doordat de gebruiker geen zicht heeft op het onderliggende mechanisme, ontbreekt elke vorm van toelichting. Een systeem dat als taak heeft om beelden te classificeren zou bijvoorbeeld aan de gebruiker kunnen aangeven dat het gelooft dat een gegeven foto een kat bevat, omdat die voor 98 procent overeenkomt met foto's van andere katten uit de database. Deze vorm van uitleg staat in schril contrast met het soort uitleg dat mensen geven. Die zouden gewoon verklaren dat het gaat om een dier met een pluizige vacht en snorharen. Het feit dat AI-systemen niet goed in staat zijn om dit soort uitleg te geven, vormt in veel gevallen een belemmering voor hun acceptatie¹⁴.

Een belangrijk punt op de agenda van de AI-gemeenschap is dan ook het streven naar AI-technieken die hun beslissingen kunnen uitleggen, de zogenaamde *explainable AI*. Mogelijk is een van de manieren om dat te bereiken het integreren van machine learning met technieken uit andere, meer traditionele deelgebieden van de AI. Het gaat hier met name om technieken die gericht zijn op het realiseren van hogere cognitieve functies, zoals redeneren, plannen en beslissen. In het vakgebied spreekt men ook wel over het integreren van symbolische met subsymbolische methoden¹⁵. Een naadloze integratie van deze Yin en Yang van de AI zou wellicht een belangrijke stap kunnen zijn richting intelligente systemen die *state-of-the-art* prestatie combineren met *high level* kennis en uitleg.

PRIORITEITEN VOOR AI-ONDERZOEK

Deze discussie geeft tevens aan dat de kunstmatige intelligentie een zeer breed vakgebied is, dat niet eenvoudig te definiëren is in termen van één specifieke techniek¹⁶. Om een heldere blik te krijgen op de fundamentele gebieden die belangrijk zijn binnen AI-onderzoek is het nuttig om in kaart te brengen waar de verschillende groepen in de AI-gemeenschap zich mee bezig houden. Voor Nederland is een dergelijke exercitie onlangs uitgevoerd door een Special Interest Group onder leiding van Koen Hindriks en mijzelf¹⁷. De resultaten zijn verwerkt in een manifest waarin een voorstel wordt gedaan voor een nationale AI-onderzoeksagenda: het *Dutch Artificial Intelligence Manifesto*. In dit manifest worden zeven fundamentele gebieden geïdentificeerd waarin Nederland sterk is, variërend van *Autonomous Agents* tot *Natural Language Processing*.

In een tijd waarin AI op het punt staat om een *disruptive technology* te worden is het essentieel dat in deze gebieden geïnvesteerd wordt. De afgelopen maanden is de zogenaamde *brain drain* van het Nederlandse AI-onderzoek veel in het nieuws geweest¹⁸: talentvolle onderzoekers kiezen steeds vaker voor een carrière in het bedrijfsleven of aan buitenlandse universiteiten. Hierdoor dreigt ons land, dat jarenlang internationaal een grote speler was, langzaam maar zeker achterop te raken. Een vergelijkbare situatie is tevens te observeren op Europees niveau. Europa is altijd sterk geweest in AI-onderzoek, maar wordt sinds een aantal jaren op verschillende fronten voorbijgestreefd door China en de Verenigde Staten, waar miljarden worden geïnvesteerd in wetenschappelijk talent.

Gelukkig bieden deze ontwikkelingen ook kansen: in China en de VS worden weliswaar grote sprongen gemaakt, maar de drijvende krachten achter het AI-onderzoek aldaar zijn substantieel anders dan in Europa. In China wordt de vooruitgang overwegend gestuurd vanuit de overheid, en in de VS komen de impulsen met name vanuit de grote technologiebedrijven. Hierdoor is de focus van het AI-onderzoek in die landen relatief eenzijdig, en wellicht niet 100 procent in overeenstemming met wat goed is voor de mensheid. Dit biedt mogelijkheden voor Europa om zich te onderscheiden met een eigen uniek profiel dat vooral ingekleurd wordt door de behoeften van de maatschappij. Inmiddels lijken steeds meer partijen zich dit te realiseren: op 7 december 2018 heeft de Europese Commissie een plan gepresenteerd waarin zij haar visie beschrijft op de toekomst van AI. In een begeleidend schrijven wordt deze visie als volgt samengevat: 'Overall, the ambition is for Europe to become the world-leading region for developing and deploying cutting-edge, ethical and secure AI, promoting a human-centric approach in the global context'¹⁹.

MAATSCHAPPELIJKE UITDAGINGEN

Kortom, binnen Europa begint men zich te realiseren dat het perspectief van de mens centraal dient te staan bij toekomstige ontwikkelingen van de AI. Maar wat houdt een *human-centric approach* precies in? Om die vraag te beantwoorden wil ik wederom verwijzen naar het eerder genoemde AI-Manifest. Daarin worden namelijk, naast de zeven fundamentele deelgebieden van de AI, ook drie multidisciplinaire, maatschappelijke uitdagingen beschreven. Deze uitdagingen kunnen worden samengevat in termen van eigenschappen die de maatschappij verlangt van moderne kunstmatige intelligentie, namelijk *Responsible*, *Explainable*, en *Socially Aware AI*.

Wat *Explainable AI* inhoudt heb ik al toegelicht aan de hand van het voorbeeld van de afbeelding van de kat. De uitdaging van de *Responsible AI* heeft te maken met het feit dat AI-systemen van nature geen besef hebben van menselijke normen en waarden. Het gevolg is dat ze beslissingen kunnen nemen die veel mensen als onethisch zouden beoordelen. Dit is met name een aandachtspunt voor toepassingen die gebaseerd zijn op het leren van data. Wanneer de aangeleverde data namelijk bepaalde vooroordelen bevatten, dan zal een AI-systeem die onmiddellijk overnemen, of zelfs uitvergroten. Een snelle zoektocht op Internet op termen als *biased AI* levert talloze voorbeelden van ongewenste gedragingen die zulke ‘slecht opgevoede AI-systemen’ kunnen vertonen, variërend van racistische voorspellingssystemen voor criminaliteit²⁰ tot seksistische chatbots²¹.

De derde maatschappelijke uitdaging betreft het ontwikkelen van *Socially-Aware AI*, ofwel AI die in staat is om te communiceren en samen te werken met mensen. Dit is het gebied waar ik mij binnen mijn eigen onderzoek voornamelijk mee bezighoud, en daarmee zijn we dan - na een lange inleiding - aangekomen bij de titel van mijn rede: Sociale kunstmatige intelligentie.

SOCIALE AI

Een van de gevolgen van de snelle ontwikkelingen binnen de AI is dat de relatie tussen mens en computer aan het veranderen is. Waar de computer in het verleden werd gezien als een slaafs apparaat dat domweg instructies van de gebruiker uitvoerde, zien we dat de scheidslijn tussen mens en computer langzaam begint te vervagen. Dit betekent niet dat de computer de mens volledig vervangt, maar wel dat beide partijen steeds nauwer samenwerken, waarbij ze idealiter elk hun eigen sterke punten benutten. Men spreekt dan ook regelmatig over ‘mens-computerteams’: complexe systemen van mensen en elektronische partners²² die hun gedrag op elkaar afstemmen. Een belangrijke consequentie van deze tendens is dat de AI-systemen van de toekomst niet alleen zelfstandig beslissingen moeten nemen, maar ook moeten kunnen interacteren en samenwerken met mensen.

Wanneer we zulke ‘sociale’ systemen willen ontwikkelen kan het nuttig zijn om ze ook te laten communiceren op een manier die voor mensen intuïtief is. Zo is het bijzonder praktisch dat wij inmiddels kunnen spreken tegen de digitale hulpjes op onze telefoons. Maar taal is niet de enige modaliteit waarmee mensen communiceren. Een groot deel van de informatie die wij onderling uitwisselen gebeurt via non-verbale communicatie zoals gezichtsuitdrukkingen en lichaamshouding. Daarom gaan sommige sociale AI-systemen nog een stap verder: ze worden uitgerust met een ‘mensachtige’ interface en met mechanismen om emoties uit te drukken en soms zelfs te herkennen in het gedrag van de gebruiker. Dit is bijvoorbeeld het geval bij sociale robots en bij intelligente karakters in computer games.

Onderzoek heeft aangetoond dat het de moeite waard kan zijn om deze sociale *agents* uit te rusten met mensachtige *gedragingen*, bijvoorbeeld omdat ze daardoor geloofwaardiger overkomen en soms zelfs effectiever zijn in het bereiken van hun doel. Binnen mijn eigen onderzoek heb ik regelmatig gebruik gemaakt van dit fenomeen. Zo heb ik aan de VU meerdere toegepaste projecten geleid die gericht waren op het inzetten van virtuele *agents* voor het trainen van communicatievaardigheden. Een goed voorbeeld betreft het promotieonderzoek van Jeroen de Man²³, later voortgezet in samenwerking met Romy Blankendaal, Charlotte Gerritsen en Marco Otte²⁴. Binnen dit onderzoek zijn de mogelijkheden onderzocht om medewerkers in het openbaar vervoer te trainen in het de-escaleren van agressie via *virtual reality*. In de ontwikkelde applicatie konden ov-medewerkers via simulatie oefenen hoe ze het beste kunnen omgaan met agressieve reizigers. De ontwikkelde trainingsapplicatie wordt sinds begin 2019 gebruikt binnen het opleidingsprogramma van de RET in Rotterdam. Een tweede voorbeeld betreft het onderzoek van Laura van der Lubbe, in samenwerking met seniorenorganisatie KBO-PCOB²⁵. Binnen dit project is een mobiele applicatie ontwikkeld voor ouderen, waarmee zij kunnen leren voorkomen dat ze slachtoffer worden van zogenaamde ‘babbeltucs’. Ook hier wordt gebruik gemaakt van gesimuleerde conversaties met virtuele *agents*, in combinatie met een module die automatisch de assertiviteit in de stem van de gebruiker analyseert. De babbeltucs-app is inmiddels gratis te downloaden, en wordt door KBO-PCOB gebruikt in hun voorlichtingsbijeenkomsten.

MULTIDISCIPLINAIR ONDERZOEK

Het moge duidelijk zijn dat dergelijke toepassingen van sociale AI grote mogelijkheden bieden. Echter, om dat potentieel optimaal te benutten is meer nodig dan alleen technologische vooruitgang. De beschikbaarheid van geavanceerde technologie op zichzelf geeft immers geen garantie voor de gewenste impact op mens en maatschappij. Het is daarom cruciaal dat we ook gedegen inzichten ontwikkelen in de *effecten* van AI op mensen en hun gedrag. Wat hiervoor nodig is voelt u waarschijnlijk al aankomen: inderdaad, *multidisciplinair onderzoek*.

Het besef dat er een sterke behoefte is aan onderzoek op het grensvlak van AI en andere disciplines is groeiende, en kan treffend worden geïllustreerd door te verwijzen naar het grote aantal leerstoelen dat de afgelopen jaren is gecreëerd op gebieden als *AI and Ethics*, *AI and Humanities* en *AI and Law*.

Wanneer we kijken naar mijn eigen onderzoek, dan is met name de connectie met de sociale wetenschappen uiterst relevant. Onderzoek naar AI die in staat is om op een natuurlijke manier met mensen te interacteren, kan vanzelfsprekend veel inspiratie putten uit hoe mensen met elkaar communiceren. Mijn nieuwe leerstoel is dan ook geworteld binnen een onderzoeksgroep die daar bij uitstek over gaat: die van de communicatiewetenschap.

COMMUNICATIEWETENSCHAP

Communicatiewetenschap is de academische discipline die onderzoekt hoe mensen met elkaar communiceren via media²⁶, ook wel *gemedieerde communicatie* genoemd. De term media duidt hierbij op een breed begrip, dat bovendien over de tijd verandert. Waar het in het verleden vooral traditionele media betrof zoals kranten en televisie, is hier de afgelopen decennia een veelheid aan 'nieuwe media' bijgekomen: eerst al de sociale media, maar inmiddels kunnen we ook spreken van een golf aan 'intelligente media'. Ik doel hierbij op digitale media die, voorzien van technieken uit de kunstmatige intelligentie, zelfstandig nieuwe mediaboodschappen produceren. Men kan hierbij denken aan online aanbevelingssystemen en gepersonaliseerde advertenties, maar bijvoorbeeld ook aan sociale robots, virtuele *agents* en chatbots.

Het zal niet als een verrassing komen dat het perspectief van de sociale AI hier naadloos bij aansluit. Waar onderzoekers vanuit de AI vooral geïnteresseerd zijn in de fundamentele technieken achter het ontwikkelen van sociale AI-systemen, houden communicatiewetenschappers zich bezig met de complementaire vraag wat deze systemen doen met mensen. Een interessante observatie daarbij is dat we, door sociale AI-systemen als media te zien, het begrip medium enigszins oprekken. Het betekent immers dat een medium niet meer altijd een 'neutraal' kanaal is voor het overbrengen van boodschappen, maar dat het ook pro-actief boodschappen kan genereren.

Als we het specifieke profiel van de onderzoeksgroep Communicatie & Media aan de Radboud Universiteit nader bestuderen, dan zien we dat deze groep zich onderscheidt met de focus op zogenaamde *positieve communicatiewetenschap*. Deze term, geïnspireerd door de positieve psychologie, is vijf jaar geleden geïntroduceerd door onze huidige programmagroepsleider, Moniek Buijzen²⁷. Positieve communicatiewetenschap verwijst naar het inzetten van media om mensen gelukkiger, gezonder en beter geïnformeerd te maken. Het initiëren van deze onderzoeksrichting was destijds een

reactie op de observatie dat het vakgebied gedomineerd werd door negatieve thema's zoals agressie en verslaving. Een goed voorbeeld van positieve communicatiewetenschap is het door Moniek geleide *MyMovez*-project, waarin wordt onderzocht hoe moderne communicatietechnologieën kunnen worden ingezet om een gezondere leefstijl onder jongeren te promoten²⁸.

Binnen de Communicatie & Media-groep worden drie thema's onderscheiden die allemaal betrekking hebben op de effecten van media op mensen, namelijk *beïnvloeden*, *informer*en en *vermaken*. Ook hier is de vergelijking met kunstmatige intelligentie, en in het bijzonder met sociale AI, moeiteloos door te zetten: een sociale robot in de klas is een voorbeeld van AI die mensen kan beïnvloeden, een chatbot op sociale media kan mensen informeren, en een AI-gestuurde tegenstander in een computer game kan mensen vermaken.

MEDIAPSYCHOLOGIE

Wanneer we de effecten van sociale AI op mensen willen bestuderen vanuit een sociaal-wetenschappelijk perspectief, dan kunnen we een aantal gevestigde theorieën als startpunt nemen. Binnen gebieden als de sociale- en de mediapsychologie wordt immers al decennialang onderzoek gedaan naar de interactie tussen mensen en media. Een belangrijk concept is daarbij het zogenaamde *antropomorfisme* – het toeschrijven van menselijke eigenschappen aan niet-menselijke entiteiten. Velen van u zijn vermoedelijk wel bekend met het experiment van Heider en Simmel uit 1944, waarbij zij proefpersonen eenvoudige filmpjes met bewegende figuren toonden²⁹. En hoewel er niet veel meer te zien was dan een cirkel en twee driehoeken, schreef men deze figuren onmiddellijk bepaalde mentale toestanden toe zoals emoties en motivaties. Kennelijk hebben we allemaal een zekere aangeboren neiging om objecten onder bepaalde omstandigheden als *agents* te zien, een gegeven waar ontwikkelaars van sociale AI dankbaar gebruik van maken.

Later gingen onderzoekers Clifford Nass en Byron Reeves nog een stap verder, door ook te kijken naar het *gedrag* dat mensen vertonen in interactie met computers. Zij introduceerden in 1996 hun *Media Equation*, die stelt dat mensen de neiging hebben om computers te behandelen alsof het echte mensen zijn³⁰. Dit onderzoek leidde in de jaren daarop tot talloze interessante studies en resultaten, zoals de bevinding dat mensen in gesprek met een computer een zekere beleefdheid aannemen. Soms vertonen mensen tijdens gesprekken met een kunstmatige *agent* zelfs nog socialer gedrag dan tegenover elkaar, zoals blijkt uit een recente studie die aantoonde dat PTSS-patiënten meer van hun persoonlijke problemen onthullen aan een virtuele therapeut dan aan een menselijke³¹.

Toch betekent dit niet dat menselijke computers in alle gevallen het gewenste effect hebben. De Japanse hoogleraar Masahori Mori toonde al in 1970 aan dat robots die erg lijken op mensen, maar er toch net even anders uitzien, vooral gevoelens van afkeer opwekken³². Hoewel het laatste woord over deze *Uncanny Valley* theorie nog niet is gezegd, geeft het wel aan dat er veel factoren zijn om rekening mee te houden bij het ontwikkelen van sociale AI. Er zijn de laatste jaren dan ook vele studies gedaan met als doel om beter te begrijpen welke aspecten van sociale AI van invloed zijn op menselijk gedrag. De opzet van dergelijke studies is meestal dat er bepaalde variabelen van de *agent* worden gemanipuleerd, zoals hun uiterlijk of hun gedragingen. Vervolgens probeert men deze factoren te relateren aan afhankelijke variabelen aan de kant van de mens, zoals de subjectieve ervaring die het bij mensen opwekt, maar ook hun fysiologische reactie, en zelfs hun gedrag.

Ook binnen onze eigen programmagroep houden verschillende collega's zich bezig met aspecten van sociale AI. Zo bestudeert Barbara Müller de (neuro)psychologische mechanismen die een rol spelen bij mens-agentinteractie, met als doel om de kwaliteit van de interactie te verbeteren. Het onderzoek van Evelien Heyselaar richt zich op het ontwikkelen van gestandaardiseerde methoden om het begrip *agency* te kunnen meten en vergelijken tussen verschillende systemen. En Paul Ketelaar onderzoekt de potentie van virtueel publiek om mensen via virtual reality te helpen van hun spreekangst af te komen. En dan heb ik nog lang niet iedereen genoemd.

Kortom, de mens-media interactie is een veelbelovend vakgebied, en meestal hoeft een *agent* niet eens in alle opzichten heel realistisch te zijn om toch geloofwaardig op mensen over te komen. Toch betekent dit niet dat we er al zijn. Hoewel we gemakkelijk kunnen worden overtuigd dat een *agent* menselijke eigenschappen bezit, en we er zelfs een soort van vriendschappelijke band mee kunnen opbouwen, kan dit effect na enige tijd ook weer vrij gemakkelijk ongedaan worden gemaakt. Een recente studie uit Tilburg illustreert dit fenomeen treffend: onderzoekers lieten proefpersonen drie weken lang communiceren met Mitsuku, momenteel de beste chatbot ter wereld. Na afloop van elke chatsessie vulden de deelnemers vragenlijsten in over hun ervaringen. In eerste instantie waren ze enthousiast over de conversaties, maar in tegenstelling tot wat de onderzoekers verwacht hadden nam dit enthousiasme over de tijd af. Op den duur ervoer men de gesprekken als oppervlakkig en afstandelijk³³.

UITDAGINGEN VOOR SOCIALE AI

Er is dus nog een hoop werk aan de winkel, en zoals eerder uitgelegd is daarvoor kennis nodig vanuit zowel de informatica als de sociale wetenschappen. Vanuit een technisch perspectief valt er bijvoorbeeld nog een grote slag te slaan op het gebied van de natuurlijke taalverwerking. De kwaliteit van chatbots is de laatste jaren al enorm verbeterd,

maar om een AI-systeem te bouwen dat een coherente conversatie kan houden over een willekeurig onderwerp zijn nog betere algoritmen nodig, die goed kunnen omgaan met context en het managen van een dialoog. Ook op het gebied van de *social signal processing* kunnen nog stappen worden gemaakt, zodat AI-systemen straks nog beter in staat zijn allerlei sociale informatie op te pikken uit het non-verbale gedrag van menselijke gesprekspartners, zoals stemgeluid, gezichtsuitdrukkingen en lichaamshouding.

Uiteindelijk zullen er ook technieken gecombineerd moeten worden. Zoals eerder aangegeven geloof ik in hybride AI-systemen, die symbolische en sub-symbolische methoden integreren. Ook voor sociale AI is dit belangrijk. Een goed voorbeeld van een aanpak voor sociale AI die verschillende methoden integreert is de *Virtual Human Toolkit* van het Institute for Creative Technologies in Californië³⁴. Dit is een modulair opgezette softwareomgeving waarmee intelligente virtuele *agents* kunnen worden ontwikkeld. De toolkit gebruikt subsymbolische methoden voor het verwerken van taal en multimodale input, en combineert deze met symbolische methoden voor onder andere het genereren van non-verbaal gedrag. Binnen mijn eigen onderzoeksgroep hanteren we een vergelijkbare aanpak. Zo ontwikkelt mijn promovendus Daniel Formolo virtuele *agents* die hun gedrag aanpassen aan hun gesprekspartner op basis van theorieën van interpersoonlijke communicatie. Ook hij gebruikt machine learning technieken voor het classificeren van multimodale input, waarna de output wordt verwerkt door een kennisgebaseerd model dat de communicatiestijl van de *agent* bepaalt³⁵.

Maar er is meer nodig dan het ontwikkelen van nieuwe technieken alleen. We willen immers ook begrijpen wat het met mensen doet om te interacteren met sociale machines. Onder welke omstandigheden accepteren wij sociale *agents* als gesprekspartners? Wat zorgt ervoor dat we een interactie met een robot als prettig, of zelfs leerzaam ervaren? Om deze vragen te beantwoorden is empirisch onderzoek nodig, iets waarin de informatica van oorsprong minder te bieden heeft. Dat brengt ons dan ook terug bij de sociale wetenschappen. Door gebruik te maken van gedegen onderzoeksmethoden kunnen we experimenten opzetten om de effecten van sociale *agents* op mensen te bestuderen. Een voorbeeld uit mijn eigen groep is het promotieonderzoek van Lenin Medeiros die middels online experimenten onderzoekt in hoeverre mensen het accepteren om emotionele steun te ontvangen van een chatbot³⁶. Ook kunnen empirische studies helpen om uit te vinden hoe we de gebruikerservaring van mensen kunnen verbeteren wanneer sociale AI-systemen nog niet perfect functioneren. Soms is het namelijk helemaal niet nodig om een algoritme verder te optimaliseren, maar kun je aan de hand van basale psychologische principes vrij gemakkelijk voor een betere ervaring zorgen, bijvoorbeeld door een robot 'sorry' te laten zeggen wanneer hij een foutje maakt.

Kortom, om het vakgebied van de sociale AI vooruit te helpen is een samenspel nodig tussen technisch en sociaalwetenschappelijk onderzoek. De informatica heeft de sociale wetenschappen nodig omdat de gemiddelde programmeur, grof gezegd maar weinig weet over de effecten van de door haar ontwikkelde systemen op mensen. En de sociale wetenschappen hebben de informatica nodig omdat experimenten met echte robots een hogere ecologische validiteit hebben dan bijvoorbeeld vignetstudies. Ik wil dan ook pleiten voor een nauwe samenwerking tussen deze twee disciplines, die wel wat weg heeft van co-evolutie: informaticaonderzoek resulteert in betere systemen die kunnen worden gebruikt als stimulusmateriaal voor gedragswetenschappelijke experimenten. De uitkomsten van de experimenten kunnen op hun beurt leiden tot *requirements* voor nog betere systemen.

ETHIEK VAN SOCIALE AI

Wat betekent dit alles voor de relatie tussen mens en computer? Op sommige gebieden, zoals schaken, accepteren we al lang dat intelligente machines er beter in zijn dan mensen. Maar als het gaat om sociale vaardigheden dan is de heersende opvatting dat robots nog ver op ons achter lopen. Toch zijn er aspecten van sociaal gedrag, zoals het herkennen van emoties, waarbij computers al alleraardigst presteren. Het is dan ook de moeite waard om te speculeren over de verdere toekomst op dit gebied. De eerder genoemde wetenschapper Harari schetst in zijn boek *Homo Deus* een extreem perspectief. Hier vergelijkt hij de sociaal-emotionele vaardigheden van een menselijke arts met die van een futuristisch AI-systeem genaamd Watson: "A human doctor recognises your emotional state by analysing external signals such as your facial expression and your tone of voice. Watson could not only analyse such external signals more accurately than a human doctor, but it could simultaneously analyse numerous internal indicators that are normally hidden from our eyes and ears. By monitoring your blood pressure, brain activities and countless other biometric data Watson could know exactly how you feel. Thanks to statistics garnered from millions of previous social encounters, Watson could then tell you precisely what you need to hear in just the right tone of voice³⁷."

Dit is een fascinerend toekomstbeeld, en mits op een verantwoorde manier gebruikt zou een dergelijk systeem beslist veel nuttige toepassingen kunnen hebben. Tegelijkertijd moeten we een belangrijk gegeven niet uit het oog verliezen: hoe slim computers ook worden aangaande vaardigheden als emotieherkenning en verbale communicatie, de voortuitgang op het gebied van computers met subjectieve ervaringen, ofwel bewustzijn, is minimaal. Of, zoals collega Pim Haselager het onlangs verwoordde: 'robots blijven ijskasten'³⁸. Wanneer we ons bezighouden met toepassingen van sociale AI moeten we ons dan ook buitengewoon goed realiseren dat die systemen hun schijnbare empathie alleen maar *simuleren*. Dat hoeft niet altijd een probleem te zijn - per slot van rekening zijn mensen ook meesters in het simuleren van empathie - maar het leidt wel tot

een ingewikkelde ethische discussie, zeker wanneer mensen op den duur vriendschappelijke, en zelfs intieme relaties ontwikkelen met kunstmatige partners. Stelt u zich eens voor hoe het is om een band op te bouwen met een wezen dat uw gedrag tot in de perfectie kan interpreteren, voorspellen, en zelfs manipuleren, maar tegelijkertijd helemaal niets om u geeft. Ik zal niet verder ingaan op de details van deze discussie, maar het moge duidelijk zijn dat het essentieel is om ons af te vragen welke toepassingen van sociale AI we wel en niet willen.

DE TOEKOMST VAN SOCIALE AI

Bijna aan het eind gekomen van mijn betoog wordt het tijd om de zaken op een rijtje te zetten. We hebben gezien dat AI wereldwijd een *disruptive technology* aan het worden is, waarin Europa en Nederland moeten investeren om niet achterop te raken. Europa kan zich daarbij onderscheiden door expliciet te focussen op wat goed is voor de *mensheid* in plaats van de overheid of de industrie. Mensen en machines werken steeds meer samen, en een belangrijke uitdaging is daarom het realiseren van sociale AI: kunstmatige intelligentie die op een natuurlijke manier kan interacteren met mensen. Om dit te bereiken is fundamenteel onderzoek nodig, maar ook multidisciplinair onderzoek op het grensvlak van de informatica en de sociale wetenschappen. Ik pleit dan ook voor een nauwere samenwerking tussen die twee disciplines, zodat ze elkaar wederzijds kunnen versterken en er bovendien aandacht kan zijn voor ethische aspecten. Met dank aan mijn nieuwe leerstoel Communicatiewetenschap en Kunstmatige Intelligentie hoop ik hieraan te kunnen bijdragen.

Mijn focus op multidisciplinair AI-onderzoek sluit bovendien mooi aan bij de missie van het recentelijk opgerichte Radboud AI-platform³⁹. Dit initiatief, dat mede tot stand is gekomen dankzij een investering van het college van bestuur heeft als doel om al het AI-onderzoek van verschillende faculteiten binnen de Radboud Universiteit en het Radboudumc te bundelen. En, toeval of niet, ook hier is ‘mensgerichte AI’ een van de speerpunten.

Hoewel de focus van mijn pleidooi lag op onderzoek, wil ik ook nog enkele woorden wijden aan multidisciplinariteit binnen het onderwijs. Immers, als we een nieuwe generatie professionals willen opleiden is het cruciaal dat die in staat zijn om zelfstandig een relatie te leggen tussen de technische en sociale aspecten van AI. Het huidige mediavertschap is extreem dynamisch, en daarom leggen we binnen de opleiding Communicatiewetenschap sterk de nadruk op het feit dat studenten de flexibiliteit moeten ontwikkelen om met nieuwe technologie om te gaan. Tegelijkertijd heeft het voor de groeiende groep AI-studenten een grote meerwaarde om iets te leren over sociaalwetenschappelijk onderzoek. Binnen mijn eigen onderwijs ben ik al geruime tijd bezig met het bouwen van dit soort bruggen, bijvoorbeeld door studenten van verschillende opleidin-

gen in teams aan projecten te laten werken. Dit soort samenwerking, tussen mensen met totaal uiteenlopende achtergronden is niet altijd makkelijk. Zeker aan het begin zijn er vaak grote verschillen, en soms zelfs bepaalde vooroordelen over de andere discipline. Wellicht heeft u daar zelf ook last van. Maar het mooie is dat, wanneer we ons goed verdiepen in onbekende gebieden, die stereotypen lang niet altijd blijken te kloppen. Zodra we ons dat realiseren komt multidisciplinaire samenwerking pas echt tot bloei.

DANKWOORD

Ter afsluiting grijp ik graag de gelegenheid aan om een aantal mensen te bedanken. Om te beginnen de leden van het college van bestuur, Han van Krieken, Daniël Wigboldus en Wilma de Koning, van het faculteitsbestuur van Sociale Wetenschappen, Michiel Kompier, Herbert Schriefers en Peer Scheepers, en de directeur van het Behavioural Science Institute, Toon Cillessen. Veel dank aan jullie allen voor het creëren van deze veelbelovende nieuwe leerstoel en voor het in mij gestelde vertrouwen.

Ook bedank ik graag al mijn fantastische collega's van de programmagroep Communicatiewetenschap onder leiding van Moniek Buijzen. Het zal voor jullie best een cultuurschok zijn geweest om plotseling zo'n exacte computerwetenschapper in jullie midden te hebben gekregen, maar dat hebben jullie niet laten merken. Vanaf mijn eerste werkdag, waarop er gelijk een prachtige bos bloemen voor mij klaar stond, voel ik mij thuis in de groep. Dat komt ook door de geweldige ondersteuning van Susanne Samuelsz die mij van begin af aan wegwijs heeft gemaakt binnen mijn nieuwe omgeving en niet te vergeten door het fijne contact met alle studenten. Met grote trots wil ik dan ook tegen jullie zeggen: 'Ich bin ein Kommunikationswissenschaftler'.

Maar hoewel ik mij inmiddels een echte communicatiewetenschapper voel, huist er uiteraard ook nog steeds een AI-er in mij. Het is dan ook bijzonder prettig dat de afdeling kunstmatige intelligentie in Nijmegen binnen hetzelfde gebouw gehuisvest is, wat het bouwen van bruggen tussen afdelingen een stuk makkelijker maakt. Ik ben Marcel van Gerven en alle andere collega's bij AI dan ook zeer erkentelijk voor alle steun en bereidheid om mee te werken aan mijn missie.

En dan mijn ex-collega's van de VU, waar ik zestien jaar gewerkt heb. In het bijzonder wil ik mijn waardering uitspreken voor mijn promotor en leermeester Jan Treur, voor alle kansen en adviezen die hij mij geboden heeft. Jan, zonder jou zou ik hier nooit gestaan hebben. Het was altijd een feest om met jou samen te werken, en aanvankelijk natuurlijk ook met Catholijn. Ik denk vaak terug aan de urenlange discussies over de kleinste details van onze computationele modellen. Op die momenten, waarop we volledig met de inhoud bezig waren, voelde ik mij op en top wetenschapper. Ook wil ik alle

andere VU-collega's bedanken voor de mooie tijd. Ik heb in zestien jaar zoveel met jullie meegemaakt, van tochtjes naar de koffiemachine tot reizen naar Australië, dat ik jullie nooit zal vergeten!

Maar bovenal wil ik mijn familie, schoonfamilie en vrienden bedanken. Lieve Joram en Mirte, broer en zus door dik en dun, wat heb ik een fantastische jeugd met jullie gehad. En stiekem hebben jullie een grote invloed gehad op de richting van mijn loopbaan: Joor door het aanwakkeren van mijn interesse in computers, en Mirt door het stimuleren van mijn sociale kant! Lieve Har, je bent de beste vader die ik mij had kunnen wensen. Dankjewel voor de heerlijke vrije opvoeding die jij en Moes mij hebben gegeven, altijd positief en stimulerend, maar nooit dwingend. Daar ben ik jullie eeuwig dankbaar voor. Lieve Moes, ook al ben je hier fysiek niet meer bij aanwezig, in gedachten blijf je altijd bij me. Ik vraag me soms af waar ik de kracht vandaan haal om zoveel ballen tegelijk in de lucht te houden in het leven. Dat weet ik nu. Dat komt door jou. Je bent onderdeel van mij geworden.

Lieve Char, we wisten allebei al lang dat jij het brein achter mijn succes bent, maar nu je een Vidi hebt binnengehaald is dat voor de buitenwereld ook bevestigd. Wat is het een voorrecht om jou in mijn leven te hebben. We kunnen samen lachen, huilen, filosoferen over de belangrijke vragen in het leven, en je helpt me nog met mijn Powerpoint-slides ook. Zoals we al regelmatig tegen elkaar zeggen: we zijn een geweldig team!

En dan het ultieme product van ons teamwork: lieve Lars en Rutger, mijn eigen sociale robotjes. Wat hebben jullie gelachen toen papa voor het eerst in deze gekke jurk liep! En wat ben ik ongelooflijk trots op jullie. Ik begon deze rede door te zeggen dat de toekomst moeilijk te voorspellen is, maar als ik naar jullie kijk dan besef ik hoe belangrijk het is om te zorgen dat die toekomst er mooi uitziet, voor jullie en voor al het andere leven op deze planeet. Laten we ons daar met zijn allen voor inzetten!

Ik heb gezegd.

NOTEN

- 1 Harari, Y.N. (2018). 21 Lessons for the 21st Century. Jonathan Cape, London.
- 2 Turing, A.M. (1936). On Computable Numbers, with an Application to the Entscheidungsproblem. Proceedings of the London Mathematical Society, 2(42), 230-65.
- 3 Wooldridge, M. (2002). An Introduction to Multiagent Systems. Wiley, New York.
- 4 Dit gebeurde tijdens het 'Dartmouth Summer Research Project on Artificial Intelligence' op Dartmouth College in New Hampshire, in de zomer van 1956.
- 5 Zie <https://www.youtube.com/watch?v=2kggRND8c7Q>.
- 6 Zie <https://www.youtube.com/watch?v=jxEo3Epc43Y>.
- 7 LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep learning. Nature 521, 436-444.
- 8 Newborn, M. (1997). Kasparov Versus Deep Blue: Computer Chess Comes of Age. Springer, Berlin.
- 9 Chouard, T. (2016). The Go Files: AI Computer Wraps Up 4-1 Victory Against Human Champion. nature.com, 15 March 2016.
- 10 Chung, J.S., Senior, A., Vinyals, O., and Zisserman, A. (2017). Lip Reading Sentences in the Wild. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
- 11 Xiong, W., Droppo, J., Huang, X., Seide, F., Seltzer, M., Stolcke, A., Yu, D., and Zweig, G. (2016). Achieving Human Parity in Conversational Speech Recognition. Technical Report MSR-TR-2016-71, Microsoft Research, October 2016.
- 12 Ciompi, F., Chung, K., van Riel, S.J., Setio, A.A.A., Gerke, P.K., Jacobs, C., Scholten, E.T., Schaefer-Prokop, C.M., Wille, M.M.W., Marchiano, A., Pastorino, U., Prokop, M., and van Ginneken, B. (2016). Towards Automatic Pulmonary Nodule Management in Lung Cancer Screening with Deep Learning. Scientific reports 7, 46479.
- 13 Een van de gevolgen hiervan is dat AI algoritmes vaak prima presteren, maar plotseling hele basale fouten kunnen maken die mensen nooit zouden maken omdat ze niet 'begrijpen' wat ze doen; zie bijvoorbeeld <http://www.evolvingai.org/fooling>. De filosoof Daniel C. Dennett spreekt in dit kader ook wel over 'Competence without comprehension'.
- 14 Miller, T. (2017). Explanation in Artificial Intelligence: Insights from the Social Sciences. arXiv:1706.07269.
- 15 Deze twee oer-klassen van de AI kunnen grofweg worden gerelateerd aan de twee systemen uit Kahneman's theorie over menselijk denken: sub-symbolische methoden vertonen een gelijkenis met Kahneman's systeem 1 in de zin dat ze snel, heuristisch en automatisch werken, en symbolische methoden lijken op systeem 2 in de zin dat ze traag, precies en berekenend te werk gaan.
- 16 Het is dan ook een misvatting om AI domweg gelijk te stellen aan Machine Learning, hetgeen tegenwoordig met name in het bedrijfsleven vaak gedaan wordt. Zoals gezegd is Machine Learning strikt genomen een van de deelgebieden van de AI, zij het een hele belangrijke.
- 17 Deze SIG bestaat uit afgevaardigden van alle nationale onderzoeksinstituten die zich met AI bezighouden. Zie http://bnvki.org/?page_id=1247.
- 18 Zie ook: 'AI voor Nederland: vergroten, versnellen, verbinden'. URL: <https://vno-ncw.nl/publicaties/ai-voor-nederland-vergroten-versnellen-en-verbinden>.

- 19 Opvallend is hier de term 'human-centric', die in veel Europese visiestukken terug te vinden is. Ook het recent gelanceerde initiatief CLAIRe (<http://claire-ai.org>), dat als doelstelling heeft om een Europees netwerk van 'Centres of Excellence in AI' op te richten, heeft een slagzin met een vergelijkbare strekking: 'Excellence across all of AI. For all of Europe. With a Human-Centred Focus'.
- 20 Zie <https://www.technologyreview.com/s/612957/predictive-policing-algorithms-ai-crime-dirty-data/>.
- 21 Zie <https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter>.
- 22 Neerinx, M.A. (2013). e-Partners op Maat. Inaugurele Rede, Technische Universiteit Delft.
- 23 Man, J. de (2016). Measuring and Modeling Negative Emotions for Virtual Training. Ph.D. Thesis. Vrije Universiteit Amsterdam, Behavioural Informatics Group.
- 24 Blankendaal, R., Gerritsen, C., Otte, M., and Bosse, T. (2018). A Virtual Reality Application for Aggression De-escalation Training in Public Transport. In: Atzmüller, M. and Duivestijn, W. (eds.), Proceedings of the 30th Benelux Conference on Artificial Intelligence, BNAIC 2018, pp. 5-20.
- 25 Lubbe, L. van der, Gerritsen, C., Formolo, D., Otte, M., and Bosse, T. (2018). A Serious Game for Training Verbal Resilience to Doorstep Scams. In: Gentile M., Allegra M., and Söbke H. (eds.), Games and Learning Alliance. GALA 2018. Lecture Notes in Computer Science, vol. 11385. Springer, Cham.
- 26 Hierbij zijn drie aspecten van communicatie van belang, namelijk de productie, de inhoud en de ontvangst van de communicatie.
- 27 Buijzen, M. (2014). Positieve Communicatiewetenschap. Inaugurele Rede, Radboud Universiteit Nijmegen.
- 28 Interessant genoeg kan er hier opnieuw een parallel worden getrokken met de Kunstmatige Intelligentie. Dit is immers ook een vakgebied waarin, zeker bij het grote publiek, vaak de nadruk ligt op de mogelijke negatieve effecten, terwijl AI eveneens een enorme potentie heeft voor wat betreft het vergroten van het menselijk welzijn.
- 29 Heider, F. and Simmel, M. (1944). An Experimental Study of Apparent Behavior. American Journal of Psychology 57:243-59.
- 30 Reeves, B. and Nass, C. (1996). The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge University Press.
- 31 Lucas, G., Rizzo, A., Gratch, J., Scherer, S., Stratou, G., Boberg, J., and Morency, L.-P. (2017). Reporting Mental Health Symptoms: Breaking Down Barriers to Care with Virtual Human Interviewers. Frontiers in Robotics and AI 4, 1-9.
- 32 Mori, M., MacDorman, K.F., and Kageki, N. (2012). The Uncanny Valley. IEEE Robotics & Automation Magazine 19(2), 98-100.
- 33 Zie <https://www.volkskrant.nl/wetenschap/mijn-chatbot-en-ik-kun-je-bevriend-raken-met-een-computer~b90d6b7d/>.
- 34 Zie <https://vhtoolkit.ict.usc.edu/>.
- 35 Formolo, D. and Bosse, T. (2017). Towards Interactive Agents that Infer Emotions from Voice and Context Information. EAI Endorsed Transactions on Creative Technologies 4(10). European Alliance for Innovation.

- 36 Medeiros, L. and Bosse, T. (2017). Testing the Acceptability of Social Support Agents in Online Communities.
In: Nguyen N., Papadopoulos G., J drzejowicz P., Trawi ski B., and Vossen G. (eds.), Computational
Collective Intelligence. ICCCI 2017. Lecture Notes in Computer Science, vol. 10448. Springer, Cham.
- 37 Harari, Y.N. (2016). Homo Deus: A Brief History of Tomorrow. Harvill Secker, London.
- 38 Zie <https://www.ru.nl/nieuws-agenda/vm/2018/oktober/pim-haselager-filmfestival-inscience-robots/>.
- 39 Zie <https://www.ru.nl/ai/>.

